

The impact of business environment and economic geography on plant-level productivity: an analysis of Indian industry^{*}

Somik V. Lall and Taye Mengistae
Development Research Group
The World Bank

June 2005

Abstract

Our analysis of manufacturing plants sampled from India's major industrial centers shows large productivity gaps across cities. The gaps partly reflect differences in agglomeration economies and in market access. However, they are also explained to a greater extent by differences in the degree of labor regulation and in the severity of power shortages. This is indication that governments can help narrow regional disparities in industrial growth by fostering the "right business environment" in locations where industry might otherwise be held back by powerful forces of economic geography. There is indeed a pattern in the data whereby geographically disadvantaged cities seem to compensate partially for their natural disadvantage by having a better business environment than more geographically advantaged locations.

Key Words: Productivity, business environment, economic geography, firm location, manufacturing, economic development, India

JEL classification: D24, L5, L6, O1, R12, R3

I. Introduction

Sub national regional disparities in industrial development are quite common in large economies. They often prompt an important policy question in developing countries -- Can governments help narrow regional gaps in industrial productivity by making the policy environment of less productive regions more business friendly?¹ Or, would the

^{*} The findings reported in this paper are those of the authors alone, and should not be attributed to the World Bank, its executive directors, or the countries they represent. The authors can be contacted at slall1@worldbank.org (Lall) and tmengistae@worldbank.org (Mengistae).

¹ We sidestep the question of whether or not this would be the right policy question. As Rice and Venables (2003) argue, regional disparities in economic performance do not necessarily signify market failure or

effects of any feasible reform in a local business environment be swamped by the forces of economic geography and path dependence? We address this issue in the context of regional gaps in manufacturing performance across India's states. Having analyzed survey data on plants sampled from the 40 main industrial cities in the country, we find that economic geography is always a major source of the variations in productivity observed across the sample.² However, the local business environment is probably an even more powerful influence. Some cities such as Bangalore and Chennai show productivity premiums primarily due to their superior geography rather than to a better business climate. At the same time business environment explains a greater share of productivity premiums or shortfalls than geography for most of the other cities. We systematically find that inferior geography appears to be compensated by better business environment in a city's industrial competitiveness. The more geographically advantaged is a state or a city, it seems, the more it can afford a "poor" business environment while maintaining competitiveness.

A significant source of the effect of geography on plant productivity is agglomeration (localization) economies, where business establishments gain productivity due to proximity to others in the same line of business. An even larger component is the effect of market access, defined as a proximity-weighted average of purchasing power or economic activity in potential trading partners of a location. For example, the city of New Delhi has a total factor productivity premium of 26 % in garments production over the city of Bhopal in the state of Madhya Pradesh on account of its economic geography. However, only 4 percentage points of this premium comes from greater localization economies. The balance of the (geographic) productivity premium reflects better access to markets for garments producers in Delhi. Similarly, the average garment producer in Mumbai has a 22 percentage point productivity advantage over the average producer in New Delhi for geographic reasons. Of this about 3 percentage points are due to

inevitably mean that their removal would improve social welfare. However, there are circumstances where reducing regional inequality is both feasible and produces welfare gains.

²The forty cities are distributed across 12 of India's 16 major states, including the most industrially advanced alongside those where the manufacturing sector is the least developed. Each of the 12 states is represented by the top 3 or 4 industrial cities in it.

Mumbai's greater localization economies, and the remaining 19 points due to its greater market access.

The productivity gap between Mumbai and Delhi underscores the importance of the distinction between the domestic and foreign components of market access. For a large economy like India domestic market access is probably the more important component for most firms. Not surprisingly, however, foreign market access explains a greater proportion of the geographic productivity premium of port cities than domestic access. Indeed, compared to Delhi, the cities of Mumbai, Chennai and Kolkata all have smaller domestic market access. The first two of these have an overall positive productivity premium over Delhi in part because their smaller domestic market access is more than compensated for by their greater foreign market access.

While the combined effect of geographic factors on productivity is quite significant, it is often smaller than the influence of two key elements of the business environment. These are (1) the degree to which labor is regulated and (2) the quality of power supply from the public grid, both of which are measured at the plant level. We measure labor regulation by whether or not the plant's management considered it to be a major or severe obstacle to the operations or growth of the plant at the time of the survey. We measure the quality of power supply by managements' estimate of the plant's output loss due to power outages in the year leading up to the survey.

These two indicators of the business environment explain a larger proportion of the productivity premium or shortfall than differences in market access or agglomeration economies. In some cities, such as Kanpur in Uttar Pradesh or Ahmedabad in Gujarat, power shortages are costlier than intrusive labor regulations. The opposite is true for Noida in Uttar Pradesh and Gurgaon in Haryana. However, in general, labor regulation and power shortages vary in the same direction, in contrast to the seeming opposition between the directions of movement of business environment as a whole and geography. We also conclude that both aspects of the business environment are equally important across the 40 cities.

An obvious policy implication of these findings is that government policies that we refer collectively as the business environment have a strong bearing on the spatial distribution of economic activities. We use the term business environment as just another name for what Dollar et al (2004) call investment climate. We define it as something broadly synonymous with institutions in the growth literature (Knack and Keefer 1995), or with what Hall and Jones (2000) call social infrastructure.³

Our focus on only two items of a possibly long list of elements of business environment is dictated by the prominence that labor deregulation and problems of the power sector have in the current policy reform literature in India. At the same time they seem to be relatively easy to measure, and show significant variation across sub national regions.⁴ Our broader result that differences in business environment are a major source of regional gaps in productivity is consistent with Dollar et al (2004) who report similar findings for international gaps in manufacturing productivity based on the World Bank's Investment Climate Surveys and the same methodology.⁵ Our findings on labor regulation are consistent with a recent paper by Besley and Burgess (2004), which also shows that regional differences in labor regulation are part of the reason why some of India's states have lagged behind others in industrial development.⁶ This finding is confirmed by Aghion et al (2004) with data disaggregated to the three-digit ISIC level

³ Accordingly we would include under it government activities relating to regulation of trade and industry and the provision of public goods that affect the cost or scale of production at the firm level by influencing factor productivity, prices or both. Apart from labor market regulations, pertinent regulatory activities include those aimed at financial markets/institutions, environmental externalities and zoning, industrial licensing, bankruptcy laws, the regulation of utilities including power, telecom or transport companies, law and contract enforcement institutions, customs regulation, and domestic and foreign trade taxes.

⁴ Many other aspects are either far more difficult to observe (e.g. the quality of law or contract enforcement institutions), or do not seem to vary a great deal sub nationally (e.g. industrial licensing and the regulation of financial markets). Besley and Burgess (2004) make the latter point with respect to labor regulation.

⁵ Our finding of association between business environment and spatial productivity gaps is also consistent with Escribano and Guasch (2004), Latin American data, Eifert et al (2005) for a number of African economies, and Haltiwanger and Schweiger (2005) on a much larger set of developing economies. Each of these papers analyzes international productivity gaps based on the World Bank Investment Climate Surveys but from perspectives slightly different from Dollar et al (2004). Dollar et al (2002) and Hallward-Driemeier et al (2003) link in country variation in productivity to business environment using the World Bank survey data for India and China respectively.

⁶ Based on an analysis of time series of aggregate (state level) data, Besley and Burgess conclude that labor law reforms that some states carried out over the period 1958-1992 to reduce compliance costs led to faster growth in manufacturing employment and productivity than in states that, either did not amend the labor law at all, or did so, but in ways that increased the cost of regulation to businesses.

and covering more recent years. It also concurs with a World Bank report on India (World Bank, 2002), which, in addition to labor regulation, identifies power shortages as a factor in economic performance gaps across states.⁷

There are two important methodological differences between the analysis of the role of labor regulation presented in this paper and that of Besley and Burgess (2004). First, Besley and Burgess (2004) measure labor regulation in terms of a state level labor dispute law index registering “pro-employer” or “pro-worker” reforms over the 34 year period covered by their study.⁸ A possible drawback in using legislative events to measure the degree of labor regulation is that it would entail suppressing possible differences across locations and regions in the implementation or enforcement of the same laws. It is also possible that firms respond differentially to the same enforcement effort due to differences in the nature and magnitude of other binding constraints. One way to minimize the bias in the estimation of local effects of labor regulation would seem to be to measure it in terms of each firm’s evaluation of how far its choices are constrained by labor laws and regulation.

The second, and more important, methodological difference between Besley and Burgess (2004) and our analysis is that we estimate the short run effects of business environment on productivity, while their analysis is of a long run nature. Our data are such that we observe the business environment in a cross section of plants over a certain year, but along side inputs and output for the same year plus the previous two years leading up to it. In the analysis we assume that the business climate indicators observed at the time of the survey also characterize a plant’s environment over the preceding two years. This enables us to use the longitudinal observations in productivity

⁷ Dollar et al (2002) also provide some evidence that differences in power shortages help to explain some of the observed regional gaps in manufacturing productivity across states.

⁸ They find that states that used their labor legislative prerogatives to introduce pro-employer amendments to India’s federal Industrial Disputes Act of 1947, had better economic performance over the period in terms of growth as well as productivity. In contrast, states that reformed the act in a “pro-worker” direction registered lower growth and lower productivity. Aghion et al (2004) also use the Besley –Burgess index in their analysis of inter-state gaps in industrial productivity and growth at the three-digit Standards of Industrial Classification Code level, basically to confirm the Besley-Burgess finding on the role of labor regulation.

to control for the technological heterogeneity of firms (as they respond to purely cross-sectional variation in business environment). The relationship between business environment and productivity that we map out in this way is a short run effect as it is conditioned on the current (or observed) spatial distribution of plants.

Profit maximizing firms would tolerate regional gaps in business environment as long as these are compensated by mobility or relocation costs, or by other sources of productivity in the current location-notably, natural geography, market access and agglomeration economies. In other words, if any two locations differed only with respect to business environment (and there were no mobility costs), all plants would eventually have (re)located away from regions of poorer business environment. This means that there would always be a “reallocation of plants” between cities and regions over any period of time in response to persistence differences in business environment.⁹ The estimates of the effect of business environment on local or regional industrial productivity we report here are short run effects as they omit (the productivity effects) of the spatial reallocation of activities.

On the other hand, being based on longitudinal analyses of aggregate (state level) data, the productivity gaps that Besley and Burgess (2004) attribute to differences in business environment are of a long run nature as they include spatial reallocation effects (as well as the short run effects). There is, however, the downside to the use of aggregate data in this context in that we cannot separately identify short run effects from aggregate data. We think that such identification is important because firms base their location choices and investment decisions on short-run productivity premiums or shortfalls, rather than on their long-run counterparts. . We also believe that the only practical way we can achieve this identification is probably estimating the effects of business environment at the plant level. At the same time, it is the long run productivity effects that would measure the true cost of possible deficiencies in a location’s business environment. Moreover, longitudinal analysis of aggregate or industry level data seems to be a far

⁹ This is not, of course, necessarily in the sense of the physical dismantling of plants to set them up elsewhere, but in that the number of plants in areas of “poor” business environment would eventually fall through closures relative to areas of relatively “good” environment, where more new plants would be setup than would be closed over the same period.

easier way for getting at this true cost than aggregating up from plant level estimates.¹⁰ We therefore would like to think of our (plant-level) approach to the analysis of the effects of business environment as a natural complement, rather than alternative, to aggregate analyses at the industry or sector level.

One advantage of estimating productivity effects of business environment at the level of the plant rather than the industry is that it enables us to identify these along side the effects of geographic variables. Simultaneous identification of the effects of institutions and economic geography would seem to be more and more difficult as we work at higher and higher levels of aggregation. It would indeed be impossible in cases where the identification of the effects of business environment has to rely on (individual) fixed effect techniques in which the effects of economic geography would have to be “differenced out”.

On the other hand, as Dollar et al (2004) point out, plant level analysis of the productivity effects of the business environment needs to address econometric problems that would not arise in the analysis of longitudinal aggregate data. One of these is the possibility of selectivity bias in least squares estimates of the effect of business environment on productivity arising from the correlation of unobservable location or firm characteristics influencing plant productivity with those affecting location decisions. There is also the possibility that the firm level incidence of business environmental bottlenecks could be endogenous even in the absence of selectivity, say, because inherently more productive firms somehow cope better with adversity in their external environment.¹¹

¹⁰ Long run effects can be identified along side short run effects based on plant level analysis. However, this requires longitudinal census data. Even then it is difficult to imagine an easier way of identifying long run effects as defined here than aggregating estimation at least at the industry level. The reason for this is that most competitive industries seem to be subject to simultaneous entry and exit, for which the spatial analogue would seem to be that new plants are observed to open in a local economy at the same time as others shutdown in the same line of business. Isolating the long run effects while keeping the analysis at the plant level then would mean identifying the effects of business environment on the productivity of the typical new entry and on that of the typical closing plant.

¹¹ This second form of endogeneity is also discussed in Dollar et al (2004) and has to be addressed in analysis based on industry aggregates as well

We have organized the rest of our paper as follows. Section 2 describes our data. Section 3 discusses econometric issues. We present estimation results in Section 4. Section 5 is summary and conclusion.

II. Data sources and measurement

Our data come primarily from the World Bank's Investment Climate Survey of India for the year 2003.¹² The survey covered 1856 manufacturing establishments sampled from 40 cities, which are the top 3 or 4 cities of the largest industrial concentration from each of 12 of India's 15 largest states. These 12 states are Andhra Pradesh, Delhi, Gujarat, Karnataka, Kerala, Haryana, Maharashtra, Madhya Pradesh, Punjab, Tamil Nadu, Uttar Pradesh, and West Bengal. Between them the 12 states account for well over 90 percent of India's industrial GDP. The 3 or 4 cities covered in each state also accounted for the bulk of manufacturing outputs of their respective states. In each city, samples were drawn exclusively from the main exporting or import competing manufacturing industries: food, textiles, garments, leather goods, drugs and pharmaceuticals, chemical, consumer electronics, electrical white goods, auto parts, fabricated metals, and machinery.

Table 1 draws up the business profile of establishments in the sample in terms of scale, age, start up conditions and productivity. The average plant has just over a hundred workers over the full sample, but with significant variation in scale across sectors, ranging from average employment size of under forty in consumer electronics to more than 250 in textiles. We are thus dealing with what is essentially a population of small to medium sized establishments, although the sample does include a sizeable number of large scale plants. The vast majority of the full sample are also owner managed, most

¹² However, we use these in conjunction with similar World Bank Surveys for other developing economies for the purpose of estimating the reference technology in terms of which the productivity indices we relate to business environment and geographic variables in the paper have been calculated. We also draw on other (secondary) sources including the Annual Survey of Industries of the Central Statistical Office of India for data on indicators of economic geography.

having been established by the current owner. Some 83 % of managers cited the desire to operate in one's state of birth was one of the major reasons for the choice of the current plant location.

Depending on its line of business, the typical plant would have been in operation at the time of the survey for between 12 years (leather goods) and 34 years (fabricated metals). During this interval it would have grown to several times of its start up size in terms of employment-ranging from just under twice the initial size in textiles to more than four times in drugs and pharmaceuticals. In spite of the seemingly high average age there seems to have been substantial turnover in the population from which our sample has been drawn. An indication of this is that a quarter of the full sample and nearly a third of those in the garments industry have been in business only for five years or less.

Plant capital intensity in terms of the book value of equipment per worker ranges from a little over US \$ 2000 in fabricated metals to about US \$ 5600 in the drugs and pharmaceuticals industry. This corresponds to a range in annual value added per worker from US \$ 2737 in fabricated metals, to US\$ 7492 in drugs and pharmaceuticals.¹³ We define value added per worker as annual sales revenue less the cost of materials and energy. The average worker earned US \$ 1087 over the full sample at the average official exchange rate of the rupee for 1999, but this also varied considerably by industry from lows of under US \$ 800 in the leather goods industry and in fabricated metals to a high of US \$ 1578 in machine tools.

(a) Business environment indices: labor regulation

The World Bank Investment Climate Survey data on India includes observations on a wide range of indicators on what we refer here as the business environment. These indicators can be classified into four broad groups: a) macro-economic stabilization; b) quality of governance including the performance of law and contract enforcement institutions; c) regulation of trade, industry, and factor markets; and d) quality of physical

¹³ All monetary figures relating to the survey are expressed in US dollars at the annual average official exchange rate of the Indian rupee for year 1999.

infrastructure, including power supply, postal and telecommunication, and transport. The consensus seems to be that there is no significant difference in terms of macroeconomic stability or the quality of law and contract enforcement institutions across states in India. On the other hand differences in business regulation and the quality of infrastructure have often been invoked as part of the reason why industry has not developed as much in some states as in others.

Within the category of business regulation, the regulation of foreign trade and customs administration come under the purview of the federal government and do not consequently vary between states. As Besley and Burgess (2004) point out, there is no difference between states in the regulation of entry and exit through licensing requirements or bankruptcy laws. This is not the case with labor regulation, which is generally believed to be more stringent in India than in most other comparable economies. Critics of the current state of labor regulation in India as a whole point out that it has unduly raised exit barriers in Indian industry and constitutes one of the major drags on its international competitiveness.¹⁴ The link between industrial exit barriers and labor regulation stems from the employment security provisions of the Industrial Disputes Act of 1947. This act sets out the rules for settlement of employment termination disputes. One of its main provisions requires establishments employing more than 100 workers to seek the permission of the state government for closure or the retrenchment of workers, which permission, critics point out, is rarely granted (Sachs et al, 1999). This is believed to have added significantly to duration of insolvency procedures in the country.

This provision of the Industrial disputes act is also criticized for having combined with other related labor laws to preclude the exploitation of economies of scale in many industries by reducing the flexibility that firms would need to adjust factor proportions to changes in product or factor market conditions. Related labor laws include the ‘service-rules’ provisions of the Industrial Employment Act of 1946 and the provisions of the

¹⁴ See, for example, Sachs, Vashney and Bajpai (1999), who blame labor market rigidity not only for keeping down the growth rate of exports low, but also for the “shockingly low” share of formal sector employment in India’s economy. The authors calculate that only 8.5 percent of India’s labor force or 27 million people are in formal sector employment of which 70 percent work for government agencies.

Contract Labor (Abolition and Regulation) Act of 1970. The Industrial Employment Act provides for the definition of job content, employee status and area of work by state law or by collective agreement, after which changes would not be made without getting the consent of all workers.¹⁵ Zagha (1999) points out that this has always made it difficult for businesses ‘to shift workers not only between plants and locations, but also between different jobs in the same plant.’ As a way out of such restrictions businesses may resort to contract workers, as per the provision of yet another law, namely, the Contract Labor Act. This law gives state governments the right to abolish contract labor in any industry in any part of the state. In states where recourse to contract labor has been more restricted as a result, keeping employment below the threshold level of 100 employees or contracting out jobs has been the only way of maintaining flexibility in the allocation of manpower.

Although labor law in India is primarily federal in origin, ample room has been provided for variation across states in effective restriction on labor market transactions by the fact that states have always been responsible for the enforcement of the same laws. They also have the power to amend some of them substantially. Based on the extent to which states have actually used this power for subsidiary legislation since the late 1950’s, Besley and Burgess (2004) have classified states into categories reflecting restrictiveness of labor regulation. Their analysis shows that states that amended laws in a “pro-labor” direction experienced slower growth on the average while those that changed laws in favor of employers grew significantly faster. In this paper we assess the link between industrial productivity and a plant-level indicator of the restrictiveness labor regulation which enables us to control for a potential gap between legislation and enforcement. The indicator we use is the plant manager’s response to the World Bank Investment Climate Survey question on whether or not managers considered labor regulation to be a major or severe obstacle to the operations or growth of their businesses. As can be seen from Table 2, there is sharp contrast in responses across states. While between one-third and 70 per cent of managers consider labor regulation as a major to severe obstacle in states such as

¹⁵ This too applies for establishments with more than 100 employees, but Zagha (1999) notes that some states have made the provisions mandatory to firms with 50 or more workers while other states have abolished the employment size limit altogether.

Tamil Nadu, Karnataka and Delhi, labor laws do not seem to draw significant complaint in the states of Uttar Pradesh and Punjab.

(b) Business environment indicators: power shortages

Disparities in power supply and transport are often cited as sources of regional gaps in economic performance in India. Regional differences in road and rail transport networks cannot be identified separately from other components of market access and should be treated accordingly as influences on productivity. When we turn to power supply the basic problem in India at the moment is one of shortage, high costs and unreliability, all reportedly stemming more from problems of transmission and distribution than of generation. There are several indicators in the World Bank Investment Climate Survey in terms of which regional differences in the gravity of the problem could be described. One of these is the frequency of outages reported by managers. According to the survey, power outages occur almost every other day for the average business in India, as compared to once every two weeks in China, and once a week in Brazil. A more accurate indicator from the point of industrial productivity seems to be, however, the output that managers estimate to lose on account of the outages. This is the indicator we use in estimating the productivity equation in the next section. According to World Bank surveys, the average manufacturer in India reports losing 8.4 percent a year in sales as opposed to under 2 percent for the average manufacturer in China or Brazil. Outages could lead to loss of sales by forcing downtime (or idle capacity) on managers. They could also cause wastage of material that would have been in-process at the point of the outage and could not be used when production subsequently resumes. Then there could be the additional cost of equipment maintenance costs directly attributable to physical damages done by outages. Regional variation in reported losses due to outages within India are shown in the first column of Table 2, losses ranging from under 4 percent in the state of Gujarat to more than 15 per cent in the state of Uttar Pradesh.

Many view the problem of power shortages in India as one of regulatory policy and ownership in the energy sector. At the root of the current supply shortage to industry

seems to be that there has been serious under-investment in transmission and distribution infrastructure over a very long time. This in turn is often linked to power generation and distribution having been a monopoly or near-monopoly of government owned enterprises operating under State Electricity Boards (SEBs). For a long time now, many SEBs have followed a deliberate policy of under pricing of electricity to households and farms, only part of which they have managed to pass on to industry through tariffs at above cost and international rates. This and the boards' growing failure to protect transmission and collect bills, has led to large financial shortfalls in relation to the needs of investment in maintenance and additional capacity. Recent efforts at attracting private investment as a solution include the opening up of generation and distribution to private capital and the unbundling of SEB's into independent commercial agencies specializing in generation, transmission or distribution only. In most states, these efforts have yet to bear fruit partly because of the reported absence of a regulatory framework in which potential investors have confidence.

(c) Indices of economic geography

Research in 'New Economic Geography' identifies market access and agglomeration economies as the two most important aspects of location specific attributes that influence cross country and sub national regional gaps in productivity (Krugman 1992; Fujita, Krugman, and Venables 1999). The purpose of introducing these variables is to examine if the benefits from agglomeration set into place a process of path dependence in terms of industrialization, and whether government policy reforms that improve the local business environment of industrially less advanced regions can counter agglomeration economies and influence the distribution of economic activities in a reasonably short time frame. Are low levels of industrial performance in some states and cities determined only by poor access to markets and lack of sufficient agglomeration economies? Or can geographically disadvantaged regions compensate their inherent deficiencies by improving their business environment to surpass those of industrially advanced regions?

Agglomeration economies are based on the technological externalities that may arise when a relatively large number of firms locate in proximity to one another.

Although these economies are external to the firm they are internal to the industries in question. They can therefore be thought of as scale economies at the level of the industry. Externalities could arise from firms sharing a pool of sector specific skills, sharing physical or institutional infrastructure for collective action or for efficient subcontracting. There is a body of evidence showing positive externalities arising from such sources. See, for example, Ciccone and Hall (1996) , and Henderson et al (1999). However, industrial concentration could lead to offsetting congestion costs by bidding up wage rates and land prices or the overcrowding transport facilities and other forms of physical infrastructure. We measure agglomeration or localization economies as the net effect of (three digit ISIC) own industry aggregate local employment on plant level productivity. Table 2 shows that, on this measure, the states of Gujarat, Delhi, Maharashtra, and Tamil Nadu would be at a clear advantage over states such as Kerala, Madhya Pradesh and West Bengal.

Market access is the proximity weighted aggregate size of the potential markets of a business or a local economy. It is sometimes divided into domestic and foreign components in the literature. For lack of a better alternative, we use as a proxy the foreign market access of a city the inverse of its distance from the nearest major port. Not surprisingly, the states that register greater agglomeration economies also have greater foreign market access on this measure. These are the states of Maharashtra, Tamil Nadu, and Gujarat. The only state that is in the top quartet in terms of agglomeration economies but does not have high foreign market access is Delhi. However, it happens to compensate for this through its greater domestic market access, which we measure using an index developed in Lall et al (2004). This is based on a gravity model whereby the market access of a location is a weighted average of the purchasing power of all other locations within India as potential trading partners, with the inverse of average travel time by road transport from the location to the trading partner as the weight attached to its purchasing power.¹⁶

¹⁶ Details of the market access index of Lall et al (2004) are given in Appendix A.

Besides Delhi other states that have relatively high domestic market access are (Table 2) Haryana, Uttar Pradesh, Punjab, Kerala and Tamil Nadu. With the exception of Tamil Nadu their advantage in domestic market access is countered by their disadvantage in terms of foreign market access, so much so that only Tamil Nadu in this group counts among the top four in terms of agglomeration economies.

(d) Measuring productivity

As already indicated we measure businesses performance in terms of productivity as we investigate the role of business environment and the economic geography of business location in its determination.¹⁷ Table 2 shows large productivity gaps between sub samples drawn from the 12 states. This is in the sense that value added per worker is, for example, 38% higher in the Karnataka sub sample than it is in the overall sample, but 22% lower in the sample from Madhya Pradesh than it is in the full sample. However, it is also clear from the same table that these gaps reflect differences in the capital or skill intensity of production or in product market structure. The regional gap in fixed assets per employee that we read from the table is indeed more pronounced than that which seen in value added per worker. There are also large gaps in the average wage rate which is likely to partly reflect cost-of-living differences across locations, but could also originate from genuine regional gaps in skills. However, we want to focus on the effects both factors would have independently of their influences on relative factor prices, that is, on their effect on total factor productivity.

Total factor productivity (TFP) is the amount by which the actual output of a plant exceeds or falls short of a counterfactual on a reference technology.¹⁸ We define

¹⁷ One reason for this choice is that a “levels” performance indicator is easier to conceptually link to cross sectional variations in one shot observations of levels of indicators of business environment or of geography than business growth or business investment. Among possible “levels” indicators technical efficiency or productivity is the more conventional measure in a context such as ours and given our data - rather than profitability-for example, partly because it is easier to measure and partly because it is easier to link to social welfare or to national economic performance.

¹⁸ The reference technology here is the average or representative production function of the population of interest. In an alternative approach the reference technology is a the frontier production function in which

separate reference technologies for each of the 11 three-digit (ISIC) industries in our data each of which we assume to be Cobb-Douglas in labor and capital inputs. We also assume that within each industry the production function is the same for all producers up to a Hicks neutral productivity shift term, A_{it} , capturing total factor productivity. Let Q_{it} be the net output of producer i in year t , $R_{it} \equiv PQ_{it}$ the corresponding revenue from sales, P is the price output, L_{it} the labor input to its production, and K_{it} the capital stock tied up with the production process over the year. To calculate TFP we estimate the parameters of

$$(1) R_{it} = A_{it}^R K_{it}^{s_R} L_{it}^{s_L} E_{it}^R$$

where $s_J^R = \frac{P_J J}{PQ}$ is the revenue share of factor J , P_J is the price of J , and E_{it}^R is an error term. If we assume that all firms have the same degree of product market power expressed in a constant price cost margin, μ , (log) total factor productivity would be given by

$$(2) \ln A = \ln A_{it}^R - (\mu - 1)(s_k^R k_{it} + s_L^R l_{it})$$

However, in the absence of knowledge of μ we observe only $\ln A_{it}^R$ and not $\ln A_{it}$.

It is therefore $\ln A_{it}^R$ that we directly relate to business environment and economic geography in the rest of the paper. This would be the same as relating $\ln A_{it}$ to the same factors as long as we control for the difference between the two indices. This we propose to do by including $s_k^R k_{it} + s_L^R l_{it}$ among the controls in the productivity equation that we estimate.¹⁹

case productivity would be decreases monotonically with technical inefficiency, which is the amount by which it falls short of a maximum defined by the frontier given the input mix.

¹⁹ See Griffith (2002) for a more detailed discussion of the problem adjusting productivity estimates for market power in a context similar to ours.

To obtain the index $\ln A_{it}^R$ itself we estimated (a log transformation of) equation (1) for each of the industries covered by the India sample, but on observations pooled from World Bank Investment Climate Survey data for seven countries including India.²⁰ Details of the estimates are given in Table 3. The first two columns of the table present OLS estimates which are, however, bound to be biased. The reason is that TFP term, a_{it}^R , is a state variable that influences the choice of inputs. As a result it must be correlated with all other right hand side variables of the equation. There are at least three recent techniques of achieving identification in this setting. These are Olley and Pakes (1996), Blundell and Bond (2000), and Levinson and Petrin (2003). We have used the Levinson-Petrin estimator because of the relative ease of its application given our data. We have too few observations on investment to use the Olley-Pakes estimator, and our panel of observations over time is too short for effective use of the Blundell-Bond approach.²¹

In the estimation of production functions labor input is often measured as the number of employees, N_{it} , of the producer. This, however, would entail ignoring the possible heterogeneity of the workforce in terms of both of skills and effort. One way of avoiding this is to take into account differences in labor productivity between producers by using the average plant-level wage rate, w_{it} , as a conversion or weight factor to define labor input as the wage bill at constant prices, that is, to set $L_{it} = w_{it} N_{it}$.

III. Econometric issues

Let us assume, for a moment, that the population of manufacturing establishments in the 40 cities and the 11 industries from which our sample is drawn is all we care about when we model the relationship between establishment level productivity, $\ln \hat{A}_{ijt}^R$, on the one hand, and business environment, x_{ijt} , and economic geography, g_{ijt} , on the other,

²⁰ The countries are Bangladesh, Brazil, China, Honduras, India, Morocco, Nicaragua, Pakistan, and Peru. The only criteria for the selection of countries was the availability of data on the industries of interest. The idea was to form as large a sample size as possible on which to base the estimation of each industry's average production function.

²¹ See Appendix B for a description of the Levinson-Petrin estimator.

where j indexes locations (or cities). To simplify notation we will treat both x and g , as scalars, although we actually estimate coefficients of two indicators of business environment and three variables of economic geography. To these we add three error components of $\ln A_{it}^R$, two of which we assume to be time invariant. Let one of these, θ_i , summarize the effect of unobservable establishment characteristics on productivity. Let the second, μ_j , be the effect of unobservable location characteristics on the same. The third, ε_{it} , is a zero mean, iid, random shocks uncorrelated with the other two and the indicators of geography and business environment. Assume also that the relationship between $\ln A^R$ and all of its determinants is linear such that

$$(3) \quad \ln \hat{A}_{ijt}^R = \alpha + \beta x_{ijt} + \gamma g_{ijt} + \delta z_{ijt} + \theta_i + \mu_j + \varepsilon_{ijt} ,$$

where $i = 1, \dots, N$, $j = 1, \dots, J$; $t = 1, \dots, T$; α, β, γ are constants, and z is a set of observed firm or establishment characteristics. We observe $\ln A^R$ and z annually for the three year period 2001-2003 so that $T = 3$. We also observe x_{ijt} only for the last of these three years, while our observation of g_{ijt} are obtained from sources other than the World Bank survey and relate to the 1990s. However, we assume that x has been constant for each firm over the three-year period of interest and that g would not have changed significantly since the late 1990s. In other words, we assume $x_{ijt} = x_{ij}$ and $g_{ijt} = g_{ij}$ so that we can re-write (3) as

$$(4) \quad \ln \hat{A}_{ijt}^R = \alpha + \beta x_{ij} + \gamma g_{ij} + \delta z_{ijt} + \theta_i + \mu_j + \varepsilon_{ijt}$$

Assuming further that both g_{ij} and z_{ijt} are exogenous to $\ln \hat{A}_{ijt}^R$, (4) can be estimated as a random effects model consistently by GLS provided that x_{ij} is also uncorrelated with θ_i and μ_j , or we can fully control for both of the latter variables. Since we have defined location at the fairly aggregate level of the city we can, in principle, control for location

fixed effects, μ_j , more or less fully. Unfortunately, we cannot say the same about controlling plant fixed effects. Given that we observe business environment and economic geography only in cross section, we cannot fully control for θ_i by differencing it out or by including establishment dummies. Yet it is not difficult to imagine scenarios where both indicators of business environment are correlated with it. It is possible, for example, that labor laws are more likely to impose binding constraints on inherently more productivity establishments, which consequently rate labor regulation as a major or severe obstacle to their growth. One reason for this could be these firms have access to more “advanced” technology that they could exploit better if the laws were different. On the other hand, inherently less productive firms, though facing the same laws would not be constrained by them not even being aware of alternative technology. Similarly, inherently more productive firms may better cope with frequent power outages through the use of more flexible processes and production schedules. They could consequently lose less in potential output than less efficient firms. We propose to correct the bias that a scenarios like these would introduce to GLS estimates of (4) by replacing x_{ij} by the corresponding city average, \bar{x}_j to estimate, instead,

$$(5) \ln \hat{A}_{ijt}^R = \alpha + \beta \bar{x}_j + \gamma_j + \delta_{ijt} + \theta_i + \mu_j + \varepsilon_{ijt}.$$

The underlying idea is that while each firm might influence the incidence of bottlenecks in its business environment, it has no control on how the average plant in its city would fare in those terms. As Dow et al (2003) note applying GLS to (5) is equivalent to the estimation of (7) by two stage least squares instrumental variable method whereby x_{ij} is replaced by its least squares predictor from the (first stage) regression

$$(6) x_{ij} = \lambda_0 + \sum_j \lambda_j D_{ij} + v_{ij}$$

where D_{ij} 's are city dummies, λ_j 's are constants, and ν_{ij} is a zero-mean, iid, random error term uncorrelated with D_{ij} 's. Only that this interpretation would not apply here since we have to include D_{ij} 's in the productivity regression in order to fully control for the unobserved location fixed effects μ_j . The latter is generally correlated with \bar{x}_j and failure to control for it would render GLS estimates of parameters of (5) inconsistent. What we should estimate rather is therefore the equation

$$(7) \ln \hat{A}_{ijt}^R = \alpha + \beta \bar{x}_j + \gamma_j + \delta_{ijt} + \left(\sum_j \phi_j D_{ij} \right) + \theta_i + \nu_{ijt}$$

where ϕ_j is a constant and ν_{ijt} is a zero mean i.i.d error term orthogonal to all the other right hand side variables. Provided that θ_i is uncorrelated with D_{ij} (or with z_{ijt}), this can be estimated consistently as a random effects model by generalized least squares.²² However, while we can always limit our firm level controls to variables that are exogenous, θ_i would be uncorrelated with D_{ij} -and hence with \bar{x}_j and g_j , only if its mean is constant across locations (or cities). It is more likely, though, that θ_i includes firm characteristics that also happen to influence plant location decisions. It is quite possible, for example, that inherently more productive plants are more likely to be found where the business environment is better, just as it is possible that more efficient firms tend to locate in geographically advantaged areas. Alternatively, the inherently more efficient could tolerate more inhospitable business environments or more adverse geography. In that case GLS parameter estimate of (7) would be subject to selectivity bias in spite of the fact that we would have fully controlled for unobserved location fixed effects.

We propose to avoid this bias by modeling productivity along with plant location decisions so that equation (7) is only the first of a two-equation simultaneous system that also includes a selection equation. In specifying the other equation, we assume that the

²² This point is also made in Dow et al (2003) in a slightly different context

location of each plant i has been chosen from a set of alternatives so as to maximize expected profits. Thus if plant i is located in city j , where expected profits are Π_{ij} , then $\Pi_{ij} > \Pi_{ik}$ for all $k \neq j$, where Π_{ik} is the profits the firm of i would have expect to earn over the life time of an alternative plant in location $k \neq j, k = 1, \dots, J$. As arguments of the profit function, wage rates and the cost of capital should depend on the very variables of economic geography and business environment included in equation (7). In addition, we assume that factor prices and location depend on unobservable factors the effect of which (on expected profits) we subsume under a location specific (time invariant) error term.

Our identifying assumption is that the magnitude of these location fixed effects in profits varies by firm type. In particular, we assume that, other things being equal, the cost of credit varies by firm type, being higher, on the average for firms that are smaller, younger, or run by newly migrant entrepreneurs. This seems to be a reasonable assumption in view of the fact that the vast majority of establishments in the sample are small to medium sized owner-managed businesses. There is considerable anecdotal and formal evidence that small businesses or new startups are rationed out of formal credit markets²³ There is also survey evidence that managers of such businesses put a premium on locating their businesses in their home towns or birth states, for which one possible reason seems to be access to credit or commercial networks. Business startup size, business age and the location of birth of entrepreneurs thus appear to be legitimate arguments of the reduced form expected profit function of new start ups or businesses contemplating a range of prospective plant locations to choose from. Of these three only business age cannot be excluded from the productivity equation justifiably. There is some evidence that the technical efficiency of plants depends on age for a variety of

²³ For example, Bigsten et al (2003) report that smaller or younger firms are less likely to get bank loans for a given level of demand for credit in a sample of firms drawn from six African countries. Levenson and Willard (2000) report that only about 6% of US businesses are rationed out of formal credit markets, but these are all small and young firms. Blanchflower, Levine and Zimmerman (1998) report that ethnic minorities are less likely to obtain formal sector loans other things being equal, and when they do typically face higher interest charges. The theory of rationing of credit (on non-price criteria) has been well established since Stiglitz and Weiss (1981), its theme being that asymmetry of information between borrowers and lenders about the risk and expected profit of projects forces lenders to resort to non price screening criteria to prevent adverse selection.

reasons. On the other hand, there does not seem to be any reason to suppose that productivity would depend on start up size once we net out scale effects, or on the location of birth of the entrepreneur except through the scale effect that this might have in as far as it could influence on access to credit.

Suppose the expected profit function can be reasonably approximated by a linear specification such that

$$(8) \quad \Pi_{ij} = a + b\bar{x}_j + cg_j + d_i\xi_j + \eta_{ij}$$

where, a , b , and c are constants; ξ_j is a location fixed effect, η_{ij} is an iid error term uncorrelated with observables and ξ_j ; and

$$(9) \quad d_i = \delta_0 + W_i'\delta_1$$

where δ_0 is a constant, δ_1 is a vector constants, and W_i is a vector of observable establishment characteristics including start up size, age and whether or not the business owner was born in j . The parameters of the profit function (8) are identified in relation to those of equation (7) by the exclusion of startup size and whether or not the entrepreneur was born in j from equation (7).²⁴

As is the common practice in the econometric literature on business location choice we assume that the error term η_{ij} has Gumbel distribution. This specifies our plant location model as conditional logit, whereby the probability that location j is chosen for plant i is given by

²⁴ Identification is possible without recourse to the exclusion of start up size and birth status of managers from the productivity equation, but only in so far as location choice is a non-linear function of observed determinants of productivity (assumed to be linear in the same determinants).

$$(10) \quad P_{ij} \equiv \Pr(\Pi_{ij} > \Pi_{ik} \forall k \neq j) = \frac{\exp(a + b\bar{x}_j + cg_j + d_i\xi_j)}{\sum_{j=1}^J \exp(a + b\bar{x}_j + cg_j + d_i\xi_j)}$$

Let $u_{ijt} = \theta_i + v_{ijt}$ so that we can re-write (7) as

$$(11) \quad \ln \hat{A}_{ijt}^R = \alpha + \beta\bar{x}_j + \gamma_{ij} + \delta_{ijt} + \left(\sum_j \phi_j D_{ij}\right) + u_{ijt}$$

Following Dubin and McFadden (1984), we can obtain consistent estimates of the parameters of (11) in one of two ways. One is to estimate it as a reduced form having replaced D_{ij} by the estimate, \hat{P}_{ij} , of P_{ij} in (10), that is, to estimate the following equation by GLS:

$$(12) \quad \ln \hat{A}_{ijt}^R = \alpha + \beta\bar{x}_j + \gamma_{ij} + \delta_{ijt} + \left(\sum_j \psi_j \hat{P}_{ij}\right) + e_{ijt}$$

where e_{ijt} is an iid random error term possibly including a plant effect that is time invariant but is nonetheless uncorrelated with all the other right hand side variables. Alternatively, we can keep the location dummies but directly control for the variation across locations of the conditional expectation of u_{ijt} . Under certain regularity assumptions, the latter is given by (Dubin and McFadden, 1984):

$$(13) \quad E(u_{ijt} \mid D_{ij}(\eta_{ij}) = j; t) = \sum_{k \neq j}^J r_k \left[\frac{P_{ik} \ln P_{ik}}{1 - P_{ik}} + \ln P_{ij} \right]$$

The selectivity bias corrected formulation of (11) is then

$$(14) \quad \ln \hat{A}_{ijt}^R = \alpha + \beta\bar{x}_j + \gamma_{ij} + \delta_{ijt} + \left(\sum_j \phi_j D_{ij}\right) + \sum_{k \neq j}^J r_k \left[\frac{\hat{P}_{ik} \ln \hat{P}_{ik}}{1 - \hat{P}_{ik}} + \ln \hat{P}_{ij} \right] + \zeta_{ijt}$$

While the estimation of equations (12) and (14) is not computationally difficult, the use of 40 city dummies with an effective sample of 1500 establishments inevitably creates a multicollinearity problem since we are measuring both business environment and indicators of economic geography at the city level. To deal with this we have defined the location dummies, D_{ij} , at the region level in terms of which we have also defined the

corresponding location probabilities \hat{P}_{ij} , while continuing to measure \bar{x}_j and g_j at the city level. This would inevitably mean that we would not be able to eliminate selectivity bias entirely. We will nonetheless know the direction in which the estimator of any of the parameters of the productivity equation is biased. We can also tell whether the estimate that we would obtain by estimating (12) and (14) with region dummies and region level correction terms is a lower bound or an upper bound of the estimate we would obtain by using city dummies and city level correction terms. This is on the assumption that selectivity bias is non-decreasing in absolute value in the level of aggregation of location.

IV. Estimation results

Preliminary estimates

Table 4 provides exact definitions and (full sample) descriptive statistics of the variables used in the estimation of the productivity equation. In Table 5 we regress total factor productivity, $\ln \hat{A}_{it}^R$, as defined by equation (1) and (2) on our indicators of business environment and economic geography with business age, line activity, region of location, and time of observation as controls. We have also included among the regressors an adjustment factor for inter firm differences in product market power. We have defined the region dummies based on the shares of states in the volume of foreign direct investment (or FDI) India attracted over the four-year period from year 2000 to year 2003. We refer to one group of states as “high FDI” states. These are the states of Maharashtra, Delhi, Tamil Nadu and Karnataka, which, together, accounted for almost 90 percent of the all India inflow of FDI over the same period (Lall and Chakravorty 2005). The second group consists of the states of Gujarat and Andhra Pradesh, which attracted significant amounts of FDI over the same period, but in volumes much lower than any of the four high FDI states. We refer to these as “moderate FDI” states. We classify the remaining 6 states as “low FDI” states, by which we mean those that did not attract any significant amounts of foreign capital over the same period. The fact that Maharashtra, Delhi, Tamil Nadu and Karnataka have attracted the bulk of FDI into India recently is consistent with

the general perception among investors that these the states have the “best” business environment in India. Responses to the World Bank Investment Climate surveys seem to confirm that this reputation holds among domestic small and medium sized firms as well. What we have grouped as moderate FDI states are also ranked higher by respondents of the World Bank survey than the remaining six states. If self-selectivity of businesses to or away from better business environment cities takes place on unobservable factors, we should observe it at the level of mobility between high FDI states and low FDI states as well.

The first panel of the table reports OLS estimates of equation (7) with location dummies aggregated into the three regions of FDI flows with moderate FDI states as the base group. The second panel presents random effects (GLS) estimates of the same specifications. We report t-values based on robust standard errors in both cases. Within each panel, specifications differ only to the extent of exclusion restrictions intended to highlight correlation among exogenous variables. The specifications that we should focus on in each panel are the first two. The difference between the two within each panel is that we adjust for differences in product market power only in one of them. The most obvious significance of the omission is that a halving of the goodness of fit of both the OLS and the GLS regressions.

Comparison of the GLS estimates with their OLS counterparts suggests the presence of heterogeneity bias: our estimate of ρ is statistically different from zero in all alternative specifications, and the GLS estimates of the effects of economic geography are as a rule significantly different from their OLS counterparts. This is setting aside the problem of selectivity bias to which we now turn, and of which there are signs already in Table 5. One sign is that, although labor regulation comes out as a statistically significant influence on productivity in the GLS panel, this is not the case with power shortages. The GLS estimates suggest that a doubling of the percentage of managers complaining about labor regulation would be associated with a four-percentage point fall in labor productivity. This seems to be a plausible figure in the light of estimates provided in other studies. On the other hand, the estimate of the coefficient of the power shortage

proxy is not statistically significant. Even if it were the magnitude the estimate would suggest that a doubling of the percentage of output lost due to power outages would reduce labor productivity by only 3 percentage points. A second indication of the presence of selectivity bias is that random effects estimates of localization economies are negative. While this could suggest congestion effects, the latter is an unlikely scenario of in the light of results previously reported for India on the subject.

Modeling location choice

As the first step for testing for this kind of bias we estimate in Table 6 the conditional logit location choice equation (10) at the level of regions rather than cities, the regions being one of high FDI states, moderate FDI states and low FDI states. It should be stressed that this is not intended to substitute for an accurate description of how businesses have sorted themselves across cities of our sample. The goal is, rather, to see if the same sorting process has significantly biased the estimates of the effects of business environment and economic geography shown in Table 5. If there is indeed a location selectivity bias in the estimates of Table 5 due to the correlation of unobservable location or business characteristics with productivity, then it is inevitable that the same bias will manifest itself in the role of business environment and economic geography on productivity gaps between high FDI states and low FDI states.²⁵

The estimates reported in Table 6 are all maximum likelihood. Because of high multicollinearity we cannot include region dummies alongside our indicators of market access and localization economies in. As a result the most general of the specifications of the Table, namely, column 3, excludes those variables. While this means that we would be unable to identify the effects of agglomeration economies and market access in location decisions, the inclusion of region fixed effects should enable us to identify the effect of business environment and firm characteristics. The sign of the fixed effects themselves partly reflects the fact that a greater proportion of the sample comes from low

²⁵ This is assuming that location choice is defined in terms of cities rather than regions.

FDI states than from high FDI states. Given the allocation of the sample between the three regions, the coefficient of the interaction term between plant age and high FDI indicates that younger plants are more likely to be found in high FDI states, meaning that more recent establishments have tended to locate in those states than in moderate or low FDI states. Given the period of establishments, larger businesses are more likely to start in high FDI states than in low FDI states as can be seen from the interaction terms of start up size and regions. The interaction terms between region and whether or not an establishment is located in the state of birth of the business owner shows that owner managed establishments are less likely to be found in the home state of the business owner in low FDI states . The coefficient of the interaction between industry groupings and regions suggests that moderate to high tech businesses are less likely to locate in low FDI states than in high FDI states.

Controlling for line of activity, time of establishment startup, scale, birth state of the business owner, and geographic characteristics of regions, the log odds ratio of a new plant locating in a region is lower where labor regulation is stronger and power shortages more severe (column 3). This is indication that excessive labor regulation and power shortages translate into lower industrial labor productivity in a city in the long run in the sense that they make firms less likely to locate plants in those cities. The corresponding short run effects are effects on the productivity of plants (conditional on location) are in Table 5, but are contaminated by the long run (or location) effects of Table 6. Our purpose in estimating equation (7) as reported in Table 6 is to identify the true short run (or plant level) effect on productivity by netting the selection effects of plant location decisions

Correcting for mobility bias

We report estimates of business environment and economic geography on plant level productivity after correcting for the selectivity bias in Tables 8 and 9. In Table 8 we do the correction by estimating equation (14) as a random effects model by GLS, that is, by including our estimate of the conditional expectation of the error term of equation (7)

computed from the location probability estimates of Table 6. A comparison of column 5 of Table 8 with those of Table 5 would seem to suggest that no selectivity bias involved in the estimates reported in Table 5. None of the selectivity terms in Table 5 are statistically different from zero, and individual coefficient estimates are not significantly different in otherwise the same specifications across the two tables. However, the reason for this outcome is the high degree of collinearity between the selectivity term and the region dummies rather than true lack of statistical significance of the former. Indeed the outcome sharply contradicts what we seen Table 9 as strong evidence for the presence of selectivity bias in Table 5. The advantage of the reduced form specification that we estimated in Table 9 is precisely that it avoids the multicollinearity problem of Table 8. We also see more plausible coefficient estimates in Table 9 than those in Table 8. This and the high values of the Hausman-Wu endogeneity test statistic reported at the bottom of column (4) to (7) lead to the conclusion that there is selectivity bias in the estimates of Table 5.²⁶

In estimating equation (12) in the first panel of Table 9, we replace the region dummies by probabilities of selection calculated from the choice probabilities of column 3 of Table 6. In the second panel of Table 9 we report the same correction for selectivity bias, but based on a multinomial selection equation in which location characteristics are assumed not to matter. This latter specification is directly estimated in Table 7, but can also be thought of as what we would obtain by imposing the restriction on equation (10) that all locations are identical in terms of observable characteristics.²⁷ In spite of the contrast between the two specifications of the selection process, the results are quite similar across the two panels in terms of the direction of selectivity bias though not in the magnitude of bias. The bias is larger when we assume that businesses self select on location characteristics as well as own characteristics. Since this is more consistent with profit maximizing behavior, and the selectivity corrected parameter estimates seem to be more

²⁶ Corresponding GLS estimators of Table 5 are used as the efficient counterpart in the computation of the Hausman-Wu test statistics reported in Table 9.

²⁷ In directly estimating this specification in Table 7 we are in effect assuming that businesses choose plant locations by some (non-profit maximizing) criterion that does not take into account location characteristics.

plausible with respect to agglomeration economies, we will focus our discussion on the first panel of Table 9.

The most important conclusion that we draw from Table 9 is probably that, in general, the effects of business environment and economic geography that we would obtain by estimating equation (7) by GLS as a random effects model would be seriously biased. Both the direction of bias and its magnitude vary by specific aspects of the business environment or geography. Thus it turns out that the uncorrected GLS estimates of the effect of labor regulation from column 7 of Table 5 is virtually identical to that in column 4 of Table 9. On the other hand, it is clear that the effect of power shortages as estimated in column 7 of Table 5 is seriously biased upwards. According to the selectivity corrected estimate of column 4 of Table 9, a doubling of the average reported loss of output due to power outages would reduce average labor productivity per plant by 22 percentage points, as opposed to reduction by only 3 percentage points according to the estimate in column 7 of Table 5. A similar comparison of column 4 of Table 9 with column 7 of Table 5 shows that estimates of the productivity effect of foreign market access that we obtain by estimating equation (7) as random effects model would not be selectivity biased, while corresponding estimates of the productivity effects of domestic market access would be upward biased by almost a factor of almost 2:1. In contrast, the uncorrected estimates of agglomeration effects in Table 5 are hugely downward biased due to selectivity.

We should perhaps stress again that the fact that we used region dummies in Tables 5, 6 and 9 in lieu of the city dummies of equations (11) to (14) and have used region, rather than city, selection probabilities, means that the estimates reported in column (4) of Table 9 would not necessarily eliminate selectivity bias entirely. However, we can tell the direction in which estimated in column 7 of Table 5 would be biased for corresponding true parameter values. More importantly, depending on the direction of this bias we can tell whether the estimates in column 4 in Table 9 are a lower bound or an upper bound of the true parameter values or to estimates that we would obtain by disaggregating the level

at which we model location choice. We can thus conclude based on Tables 5 and 9 that, there is no indication that the estimates of the effect of labor regulation or of foreign market access on productivity obtained by applying GLS to equation (7) is selectivity biased. On the other hand our estimates of agglomeration economies obtained in the same way are seriously biased downwards while the corresponding estimates of the productivity effects of power shortages and domestic market access are seriously biased upwards. Since the direction of bias is unlikely to change as we disaggregate the level of location choice, we conclude that our estimates of the effects of labor regulation and foreign market access as reported in column (4) of Table 9 are consistent. On the other hand our estimates of the effect of power shortages reported in the same column should be an upper bound for the true value as must be the effect of domestic market access. This means that our estimates of agglomeration economies and of the effect of power shortages on productivity should, if anything, understate the true effect, while we could be overestimating the effect of domestic market access.

With this caveat in mind, the key finding summed up in column 4 of Table 9 is that both business environment and economic geography are powerful influences on plant level productivity. Concerning the influence of geography we find that a doubling of a city's domestic market access relative to the all India maximum would almost triple labor productivity in the average plant in that city. Secondly, a doubling of the distance of the city from the nearest port relative to the all India maximum would reduce labor productivity in the average plant by 10%. Third, a doubling of the own industry employment size in the city relative to the all India maximum would increase labor productivity in the average plant by almost 40%. One piece of evidence that business environment is a powerful influence on plant productivity is that a doubling of the percentage of businesses in the city that feel seriously constrained by labor regulation would be associated with 4% decline in labor productivity. A second is that a doubling of businesses estimates of percentage of sales lost due to outages would be associated with a 22% loss in labor productivity.

Accounting for inter-city productivity gaps

What do these results tell us in terms of the relative importance of business environment and economic geography as determinants of plant level productivity? And what can we infer from them about the relative importance of individual elements of business environment or of economic geography? Because of differences in units of measurement, we cannot say anything about either of these questions simply by looking at coefficient estimates in Table 9. We can, however, estimate the fraction of the observed gap between cities and regions that can be attributed to business environment or to economic geography or to any components thereof based on the coefficients. In Figure 1 we show how the productivity gap among garments producers that we observe in our data across cities are divided into a component due to differences in business environment and one due to differences in economic geography. In the chart the average log of total productivity for the state of Delhi has been normalized to zero. One obvious result from the figure is that the business environment component is larger. It is true that garments producers in cities such as Bangalore and Chennai owe their greater productivity to the more favorable economic geography of their location rather than to their business environment being better. However, it turns out that business environment accounts for the greater share of the productivity premium or shortfall of most other cities.

A second striking pattern in Figure 1 is that the cities where economic geography is less favorable tend to have a better business environment. It is as if the typical city compensates for adverse economic geography by improving its business climate or locations with more favorable geography can afford to have a poorer business environment. There are notable exceptions to this pattern including the cities of Mumbai, Chennai, Surat and Comibatore, all which seem to have a productivity premium over Delhi on account both of better economic geography and of better business environment.

Figure 2 shows the relative strength of the two components of economic geography. The chart leaves little doubt that market access is by far the more important element in terms of explaining the productivity gaps we observe across the forty cities.

However, it is also clear that differences in agglomeration economies typically account for fairly large fractions of those gaps. For example, of the 26 % geography related productivity shortfall of the average garments producers in the city of Bhopal (in the state of Madhya Pradesh) relatively to Delhi, 4 percentage points are due to the fact that there is a larger cluster of garment producers in New Delhi. The balance of shortfall reflects that Delhi has far greater market access than Bhopal. Similarly, 3 percentage points of the 22 percentage point productivity premium of the average garment producer in Mumbai over its counterpart in Delhi are due to the greater concentration of garment production in Mumbai. The remaining 19 percentage points reflect Mumbai's greater market access.

The productivity gap between plants in Mumbai and those in Delhi underscores the importance of the distinction between the domestic and foreign components of market access. It also illustrates that, although domestic market access is probably the more important of the two components for an economy as large as India's, foreign market access could be more important for coastal regions. This is what we see in Figure 3, where most cities are shown to "lose out" in productivity to Delhi simply because Delhi has far greater domestic market access. At the same time we see that garments producers in several major port cities including Mumbai and Chennai have a productivity advantage over their counterparts in Delhi in spite of their lower domestic market access. The reason for this seems to be that as port cities, greater foreign market access more than makes up for that deficiency.

In Figure 4 we bring together the various components of business environment and economic geography with overall inter-city productivity gaps, again among garments producers. The chart highlights further two points already made with the other charts. One of these is that differences in business environment explain more of the variation in productivity across cities than economic geography does. The second is that economic geography and business environment tend to move in opposing directions in accounting for inter-city gaps in the sense that cities that have large shortfalls on account of less favorable economic geography tend to have a better business environment. In addition, Figure 4 shows that business environment and economic geography together explain a far

larger proportion of inter-city variation in average business productivity than is suggested by the relatively low measures of goodness of fit of the model estimated in Table 9 to firm level observations. In particular, the proportion of inter-city gaps in productivity due to differences in business environment is not only larger than that attributable to geography, but it is a high proportion of the overall productivity gap. In most cases the productivity gap attributable to business environment is more than a third of the total.

We can assess the relative importance of the two components of business environment with the help Figure 5. Power shortages are clearly the more costly deficiency in cities such as Kanpur in Uttar Pradesh and Ahmedabad in Gujarat. However, this is not the case with cities of Noida (Uttar Pradesh) and Gurgaon (Haryana), where differences in labor regulation explain a greater proportion of productivity gaps compared to Delhi. It indeed seems safe to conclude that both aspects of the business environment are equally important sources of gaps when considered across the 40 cities as a whole. It also appears that, in general, labor regulation and power shortages vary in the same direction, in contrast to the seeming opposition between the directions of movement of economic geography and business environment as a whole.

V. Conclusion

In this paper, we have analyzed the productivity of manufacturing plants sampled from 40 of India's largest industrial cities. The analysis shows that there are large productivity gaps across locations which, to a large extent, we explain by differences in two attributes -- economic geography, and local business environment.

This indicates that governments can help bridge regional disparities in industrial development in India by inducing the "right business environment" in locations where economic geography is not as favorable as in industrially advanced cities and regions. There is of course nothing in our analysis that shows that regional disparity in industrial development is necessarily bad in terms of aggregate efficiency. To the extent that bridging regional gaps in industrial performance is desirable, however, our results suggest that governments can help achieve this by improving the business environment of

locations despite the presence of strong forces of economic geography that tend to generate or magnify the gaps.

India is not the only large developing economy where regional disparities in industrial development constitute an active policy issue. In various forms and to different degrees regional equity in industrial development seems to be a policy issue in countries such as China and Brazil as well. We therefore hope that our results would inform prospective studies in that broader context. To the extent that regional inequality in industrial development is a legitimate policy concern, we hope the paper contributes to better understanding of the business environment as an instrument of policy intervention. The notion seems too common in development circles that better performing locations or regions must have better business environment on the average. Our paper suggests that this is not necessarily true unless one defines business environment widely enough to subsume all aspects of the economic environment of producers including the region's economic geography. This would certainly make the concept of business environment tautological and redundant. If we define it to include only those aspects of the economic environment that government policy can influence in the medium to the short run, then it is not necessarily the case that the most productive locations have the best policy environment. The case for improving the business environment of relatively backward regions through policy reforms need not thus be that those policies are necessarily worse than those in more advanced locations. It can be that, in the presence of sources of cost advantages no one controls, it makes sense for governments to compensate for them by acting on factors that they can influence.

Our findings should be qualified by a major data limitation that we have pointed out in the paper. This is that our data come from the most industrially advanced cities in India. In terms of regions, each state is represented in the World Bank Investment Climate Survey by the top three or four industrial cities in that state. We are therefore comparing locations across the top end of the size distribution of cities and urban centers. It is possible that as we move down the distribution the relative importance of geography and business environment shifts from what is reported here. It is difficult to say if these

findings are representative of smaller cities. However, the disparity among the forty cities analyzed here is a large component of the spatial inequality in industrial activity in India at the moment, and is probably what many people have largely in mind when the issue of regional (as opposed to local) inequality is raised in the policy literature.

References:

- Aghion, P., R. Burgess, S. Redding, and F. Zilibotti. 2003, "The Unequal Effects of Liberalization: Theory and Evidence from India", Processed, London School of Economics, October.
- Bastos, F., and J. Nasir. 2004. "Productivity and Investment Climate: What matters Most?" World Bank Policy Research Working Paper No. 3325.
- Besley, T. and R. Burgess, 2004, Can labor regulation hinder economic performance? Evidence from India, *Quarterly Journal of Economics* 119(1), 91-134, February.
- Bigsten, A., P. Collier, S. Dercon, M. Fafchamps, B. Gauthier, J. Gunning, A. Oduro, R. Oostendrop, C. Pattillo, M. Sodorbom, F. Teal, A. Zeufack. 2003. "Credit Constraints in Manufacturing Enterprises in Africa," *Journal of African Economies*, vol. 12 (1), pp. 2003.
- Blanchflower, D. P. Levine, and D. Zimmerman. 1998. "Discrimination in the small business credit market," NBER Working Paper # 6840, December..
- Ciccone, A. and R. Hall. 1996. "Productivity and the Density of Economic Activity," *American Economic Review*, vol. 86, no. 1, pp. 54-70.
- Dollar, D., I. Giuseppe, and T. Mengistae, 2002, "Investment Climate and Economic Performance: Some Firm Level Evidence from India" Working Paper No. 143, Center for Economic Research and Policy Reform, Stanford University.
- Dollar, D., M. Hallward-Driemeier, and T. Mengistae, 2004, "Investment Climate and Firm Performance: in Developing Economies," Processed, World Bank, February
- Dubin, M., and D. McFadden, 1984. "An Econometric Analysis of Residential Electric Appliances Holdings and Consumption," *Econometrica*, vol. 52, pp. 345-362.
- Eifert, B., A. Gelb, and R. Ramachandran. 2005. "Business Environment and Comparative Advantage in Africa: Evidence from the Investment Climate Data," mimeo, World Bank Group, Washington DC, January.
- Escribano, A. and J. Guasch. 2004. "Assessing the Impact of Investment Climate on Productivity Using Firm Level Data: Methodology and the cases of Guatemala, Honduras and Nicaragua," mimeo, World Bank Group, Washington, DC., November.

- Fujita, M, P. Krugman, and A. Venables. 1999. *The Spatial Economy: Cities, Regions and International Trade*. The MIT Press: Cambridge, MA..
- Hall, Robert E., and Charles Jones 1999. “Why Do Some Countries Produce So Much More Output per Worker than Others?” *Quarterly Journal of Economics*, volume 114: (February): 83-116.
- Hallward-Driemeier, M. S. Wallsten and L. Xu, 2003, “The Investment Climate and the Firm: Firm Level Evidence from China,” World Bank Policy Research Working Paper No. 3003.
- Haltiwanger, J. and H. Schweiger. 2005. “Allocative Efficiency and Business Climate,” mimeo, University of Maryland, January.
- Henderson, V., A. Kuncoro and M. Turner (1995), “Industrial development in cities. *Journal of Political Economy*”, 103, 1067-1090.
- Knack, Steven and Philip Keefer.1995. “Institutions and Economic Performance: Cross Country Tests Using Alternative Measures.” *Economics and Politics* 7:207-227
- Krugman, P.1991. “Increasing Returns and Economic Geography.” *Journal of Political Economy*, vol. 99, pp.483-499.
- Lall, S.V., Z. Shalizi and U. Deichmann. (2004). “Agglomeration Economies and Productivity in Indian Industry.” *Journal of Development Economics*, 73, 2, 643-673, 2004.
- Lall, S.V. and S. Chakravorty .2005. “Industrial Location and Spatial Inequality: Theory and Evidence from India”. *Review of Development Economics*, 9, 47-68, 2005.
- Levenson, A. and K. Willard. 2000. “Do Firms Get the Financing they Want? Measuring Credit Rationing Experienced by Small Businesses in the U.S., *Small Business Economics*, vol. 14, pp. 83-94.
- Levinsohn, James and Amil Petrin 2003. “Estimating Production Functions Using Inputs to Control for Unobservables.” *Review of Economic Studies*, vol . 70 (2), pp. 317-342
- Olley, Steven and Ariel Pakes 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry.” *Econometrica*, 64: 1263-1297

- Rice, P. and A. Venables. 2003. 'Equilibrium Regional Disparities: Theory and British Evidence,' *Regional Studies*, vol. 37, no. 6& 7, pp. 675-686
- Sachs, J., Varshney, A. Bajpai, N. (eds) (1999). *India in the Era of Economic Reforms*. New Delhi: Oxford University Press
- Stiglitz, J. and A. Weiss. 1981. "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, vol.71, no. 3, pp. 393-410.
- Zagha, R. 1999. "Labor and India's Economic Reforms." In Sachs et al. (eds), *India in the Era of Economic Reforms*. New Delhi: Oxford University Press.

Appendix A: Measuring Domestic Market Access

In computing domestic market access, Lall et al (2004) use the negative exponential weighting formula to obtain the index: $I_i^{ne} = \sum_j S_j e^{(-d_{ij}^b / 2a^2)}$, where I_i^{ne} is the market access index of location i , S_j is the purchasing power of potential trading partner (location) j , d_{ij} is the distance between i and j , b is a constant describing the rate at which interaction between locations i and j decreases with the distance between them, and a is a constant point of inflection of the negative exponential function. In actual computation Lall et al (2004) proxy S_j by the population of j , while they measure d_{ij} by the estimated average time it takes to travel between i and j , on the Indian road network. The advantage of using average travel times rather than distance is that travel time would take into account the influence of road and vehicle quality as well as topography on effective interaction between locations. Their market access index is therefore positively correlated with road transport as part of the physical infrastructure on which business in i operate. To the extent that the purchasing power of i itself enters the computation of I_i^{ne} , the index should also be a good proxy for the urbanization economies that businesses located in i could benefit from. To the extent that the market access variable reflects the quality of transport infrastructure it should be positively correlated with factor productivity. To the extent it reflects urbanization economies of or access to external markets, it should influence the cost of inputs (or supplier access) as well as enabling firms to enjoy scale economies by increasing the demand for firms products.

Appendix B: The Levinshon-Petrin Estimator

The Levinshon-Petrin identification strategy is based on the assumption that the demand for intermediate inputs depends on $\ln A_{it}^R$ and another state variable, namely, capital stock, k_{it} , that is, $m_{it} = m_t(k_{it}, \ln A_{it}^R)$. Levinshon and Petrin (2003) show that provided that (a) technology is such that the production function is twice differentiable and the cross partial derivatives of m_{it} and $\ln A_{it}^R$ exist, (b) all firms are price takers, and (c) current investment does not influence current productivity, m_{it} is monotonically increasing in $\ln A_{it}^R$ so that the intermediate input demand function can be inverted to express $\ln A_{it}^R$ as a function of capital stock and intermediate input demand:

$\ln A_{it}^R = a_t(k_{it}, m_{it})$. If we further assume that productivity is a first-order Markov process, that is, if $\ln A_{it}^R = E[\ln A_{it}^R | \ln A_{it}^R] + e_{it}^R$, where $E(k_{it} e_{it}^R) = 0$, the influence of e_{it}^R on input demands can be controlled for fully in the estimation of the production function so as to avoid the simultaneity bias OLS estimates of factor shares are subject to. Levinshon and Petrin (2003) propose a two stage estimation procedure in which consistent estimates all factor share parameters of the production function can be obtained.

The first of these stages involves the consistent estimation of the share of non-state input variables by applying OLS to the log transformation of equation (3), but having augmented it by including a third order polynomial approximation in k_{it} and m_{it} of the unknown a_t function as a term. Parameter estimates resulting from this step are then used to predict $\ln A_{it}^R$. In the second stage a consistent estimate for the share of capital is obtained by minimizing the excess of capital over the shares of all factor inputs and the mean of the predictions of $\ln A_{it}^R$ obtained from the first step.

Table 1 : Productivity profiles and business characteristics across sector

sector	Value added per worker (US \$)	Fixed assets per worker (US \$)	Average annual wages per worker (US \$)	Average Number of workers	Average age of business (years)	Number of employees at startup	Proportion of businesses in state of birth of owners	Proportion of businesses aged five years or less
Food & bev	5938	3741	1114	86	16	32	0.81	0.30
Textile	5123	4440	1292	258	19	134	0.76	0.19
Garments	3624	2006	852	52	20	25	0.86	0.31
Leather	4751	2702	762	67	12	18	0.83	0.29
Chemicals	5113	4600	1013	101	14	25	0.82	0.23
Pharma	7492	5658	1374	216	32	49	0.88	0.19
Metal works	2737	2045	749	43	34	14	0.82	0.25
Consumer electronics	4520	2616	1067	38	26	15	0.90	0.24
Electrical goods	4311	2674	1159	48	13	16	0.86	0.21
Auto components	3683	3064	1056	48	14	21	0.79	0.24
Machine tools	6748	3875	1578	84	16	27	0.93	0.16
All sectors	4783	3433	1087	101	20	39	0.83	0.24

Table 2: Business environment, economic geography and labor productivity across states

State	Log difference with all India average per establishment in				Business environment indicator		Geographic variables		
	Annual value added per worker	Fixed assets per worker	Annual wage bill per worker	Estimated price cost margin	Output lost to power outages (%)	Businesses constrained by labor regulation (%)	Domestic market access index	Log of distance from nearest port	Own industry local employment
Andra Pradesh	0.19	0.11	-0.05	0.02	5.29	2.46	6.88	5.92	4.39
Delhi	0.11	0.31	0.15	0.00	11.12	68.53	6.99	6.84	4.52
Gujarat	0.38	0.17	0.02	0.05	3.51	7.04	6.87	4.26	4.54
Karnataka	0.10	0.25	0.00	0.07	12.63	50.07	6.89	4.42	4.45
Kerala	0.04	0.13	-0.25	0.06	9.47	6.04	6.93	2.99	4.35
Maharashtra	-0.05	-0.42	0.16	-0.03	6.41	17.77	6.88	3.81	4.51
Punjab	0.09	0.33	0.17	-0.01	10.40	2.75	6.93	6.96	4.43
Tamil Nadu	-0.04	-0.05	-0.15	-0.08	7.20	28.74	6.91	3.70	4.49
West Bengal	-0.13	-0.45	-0.20	0.03	6.00	9.09	6.88	4.04	4.30
Uttar Pradesh	-0.01	0.24	0.03	0.00	15.08	2.12	6.97	6.84	4.43
Haryana	-0.10	0.28	0.21	-0.08	8.62	22.27	6.97	6.84	4.41
Madhya Pradesh	-0.22	-0.38	-0.24	-0.01	10.12	2.07	6.86	6.18	4.36

**Table 3: OLS and Levinsohn-Petrin Estiamtes of Cobb-Douglas Specification of Production Function
by sector**

Dependent variables= Log (annual value added)

	OLS estimates		Levinsohn-Petrin estimates		Observations	Number of Firms
	Log (fixed assets)	Log (wage bill)	Log (fixed assets)	Log (wage bill)		
All firms	0.266 (57.88)**	0.832 (175.91)**	0.138 (5.20)**	0.551 (43.36)**	18842	7668
All nine sectors	0.280 (56.18)**	0.818 (158.98)**	0.133 (5.36)**	0.546 (42.93)**	16516	6662
Food /beverages	0.270 (17.01)**	0.836 (49.96)**	0.002 (0.02)	0.570 (18.62)**	2009	839
Textiles	0.302 (24.80)**	0.788 (60.02)**	0.172 (3.34)**	0.487 (24.04)**	3084	1216
Garments	0.305 (31.55)**	0.793 (83.45)**	0.114 (3.05)**	0.543 (26.35)**	4190	1686
Leather goods	0.276 (16.09)**	0.840 (50.11)**	0.203 (3.16)**	0.514 (10.43)**	1095	439
Chemicals, Drugs and pharmaceuticals	0.297 (20.91)**	0.801 (52.74)**	0.251 (3.17)**	0.502 (15.37)**	2666	1126
Electrical goods/equipment	0.177 (10.60)**	0.920 (52.95)**	0.085 (0.86)	0.664 (12.00)**	1322	537
Auto parts	0.269 (12.93)**	0.820 (37.54)**	0.074 (0.96)	0.558 (6.75)**	1057	391

Absolute value of z-statistics in parentheses

* significant at 5% level; ** significant at 1% level

Table 4: Descriptive statistics

Variable	Definition	Mean	Std. Dev.
<i>Productivity variables:</i>			
Value added	Log of annual value added per worker (US \$)	7.93	1.03
Fixed assets	Log end-of-fiscal year book value of fixed assets (in US \$)	7.44	1.16
Average wage rate	Log of annual wage bill per employee (US \$)	6.68	0.73
Employment	Log end-of-fiscal year number of employees	3.24	1.29
<i>Business characteristics:</i>			
Business age	Log number of years since the establishment was set up	2.48	0.81
Start up size	Number of employee's at the startup of the establishment	2.49	1.09
Entrepreneurial roots located	Dummy=1 if the business is located in the birth state of the current owner	0.83	0.38
<i>Business environment variables</i>			
Output loss to power outage (%)	Log of city average of estimated % of annual output lost due to power outages	2.06	0.51
Constrained by labor regulation (%)	Log of proportion of businesses held back by labor regulation in the city	1.94	1.49
<i>Geographic variables:</i>			
Foreign market access	Log distance of city from the nearest major port	3.17	1.42
Domestic market access	Log of ratio of city's domestic market access index to the all India maximum	6.91	0.05
Own industry local size	Log of own industry employment in the city	4.44	0.09
<i>Sector:</i>			
Food	Dummy=1 if in the food and beverages industry	0.11	
Textiles	Dummy=1 if in textiles	0.13	
Garments	Dummy=1 if in garments	0.14	
Leather goods	Dummy=1 if in leather goods	0.04	
Drugs and pharmaceutical	Dummy=1 if in drugs and pharmaceuticals	0.10	
Chemicals	Dummy=1 if in chemicals	0.11	
Consumer electronics	Dummy=1 if in consumer electronics	0.08	
Electrical white goods	Dummy=1 if in electrical white goods	0.07	
Auto parts	Dummy=1 if in the auto components industry	0.13	
Fabricated metals	Dummy=1 if in fabricated metals	0.05	
Machine tools	Dummy=1 if in machine tools	0.04	

Table 5: OLS and GLS Estimates of TFP Equation Uncorrected for selectivity bias

Dependent variable Log TFP (adjusted for skills)

	OLS						Random effects (GLS)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Adjustment for market power	1.842 (50.28)**		1.839 (50.43)**	1.843 (50.26)**	1.838 (50.37)**	1.860 (50.07)**	1.524 (51.41)**		1.523 (51.44)**	1.524 (51.39)**	1.522 (51.42)**	1.526 (51.32)**
Foreign market access	-0.120 (10.05)**	-0.147 (9.54)**	-0.122 (10.40)**	-0.111 (9.64)**	-0.114 (9.97)**		-0.113 (6.08)**	-0.143 (5.89)**	-0.115 (6.30)**	-0.103 (5.71)**	-0.105 (5.92)**	
Domestic market access	2.887 (7.21)**	3.311 (6.34)**	2.845 (7.15)**	2.602 (6.70)**	2.531 (6.59)**		3.155 (4.79)**	3.540 (4.13)**	3.106 (4.75)**	2.805 (4.39)**	2.719 (4.31)**	
Own industry local size	-0.166 (0.95)	-0.010 (0.04)	-0.148 (0.86)	-0.102 (0.59)	-0.074 (0.43)		-0.128 (0.46)	0.035 (0.10)	-0.110 (0.40)	-0.044 (0.16)	-0.016 (0.06)	
Constrained by labor regulation (%)	-0.033 (2.84)**	-0.032 (2.12)*	-0.034 (2.97)**			-0.005 (0.48)	-0.040 (2.14)*	-0.045 (1.85)	-0.041 (2.23)*			-0.013 (0.74)
Output lost to power outages (%)	-0.032 (0.94)	-0.043 (1.28)				-0.076 (2.33)*	-0.031 (0.61)	0.098 (1.47)		-0.044 (0.87)		-0.067 (1.35)
Business age	0.158 (10.13)**	0.182 (8.94)**	0.159 (10.22)**	0.154 (9.93)**	0.156 (10.03)**	0.160 (10.15)**	0.161 (6.91)**	0.199 (6.57)**	0.162 (6.96)**	0.157 (6.76)**	0.158 (6.82)**	0.164 (7.00)**
High FDI State	-0.478 (9.61)**	-0.579 (10.05)**	-0.500 (11.39)**	-0.518 (10.87)**	-0.551 (13.60)**	-0.344 (7.03)**	-0.474 (6.22)**	-0.652 (6.58)**	-0.495 (7.25)**	-0.522 (7.15)**	-0.554 (8.79)**	-0.338 (4.52)**
Low FDI State	-0.529 (10.32)**	-0.581 (9.82)**	-0.552 (12.27)**	-0.502 (9.95)**	-0.532 (11.94)**	-0.412 (8.72)**	-0.557 (6.96)**	-0.702 (6.75)**	-0.580 (8.17)**	-0.526 (6.68)**	-0.557 (7.92)**	-0.434 (5.97)**
Constant	-23.991 (9.42)**	-18.441 (5.54)**	-23.807 (9.38)**	-22.368 (9.01)**	-22.031 (8.92)**	-5.328 (27.21)**	-24.479 (5.86)**	-20.431 (3.76)**	-24.254 (5.83)**	-22.504 (5.52)**	-22.091 (5.46)**	-3.724 (20.51)**
Industry dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3525	3525	3525	3525	3525	3531	3525	3525	3525	3525	3525	3525
Establishments							1431	1431	1431	1431	1431	1431
R-squared	0.60	0.31	0.60	0.60	0.60	0.59	0.6	0.31	0.6	0.59	0.6	0.58
Rho							0.88	0.88	0.88	0.88	0.88	0.89

Absolute value of t statistics in parentheses

* significant at 5%; ** significant at 1%

Table 6: Maximum likelihood estimation of conditional logit specification of location choice

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Interaction of medium to high tech industries**** with:</i>						
High FDI States			-0.903 (5.00)**	-1.232 (7.08)**	-1.005 (5.71)**	-0.923 (5.02)**
Low FDI States			-1.056 (5.72)**	-1.231 (6.94)**	-1.109 (6.17)**	-0.999 (5.27)**
<i>Interaction of "entrepreneurial roots located" businesses with</i>						
High FDI States			-0.109 (0.33)	-0.369 (1.14)	-0.178 (0.54)	-0.012 (0.04)
Low FDI States			-1.874 (6.05)**	-1.797 (5.93)**	-1.850 (6.06)**	-1.699 (5.21)**
<i>Interaction of startup size with</i>						
High FDI States			-0.161 (2.46)*	-0.240 (3.86)**	-0.184 (2.91)**	-0.159 (2.39)*
Low FDI States			-0.351 (4.98)**	-0.367 (5.46)**	-0.355 (5.19)**	-0.326 (4.51)**
<i>Interaction of business age with:</i>						
High FDI States	-0.362 (3.56)**		-0.269 (2.38)*	-0.330 (3.10)**	-0.289 (2.65)**	-0.282 (2.45)*
Low FDI States	-0.217 (2.20)*		-0.015 (0.13)	-0.082 (0.74)	-0.035 (0.31)	-0.017 (0.15)
Constrained by labor regulation (%)			-0.344 (4.09)**	-0.645 (8.13)**	-0.913 (10.41)**	
Output lost to power outage (%)			-2.058 (10.31)**			-3.995 (14.44)**
Domestic market access		-2.913 (0.74)		5.000 (1.59)		
Own industry local size		-3.650 (2.85)**			8.902 (6.44)**	-16.983 (8.60)**
High FDI States	1.613 (5.92)**		2.746 (5.73)**	3.206 (6.85)**	2.644 (5.60)**	3.033 (6.12)**
Low FDI States	1.373 (5.12)**		4.151 (8.69)**	4.186 (9.05)**	4.035 (8.64)**	4.068 (8.24)**
Log likelihood	-1541	-1619	-1291	-1348	-1328	-1256
Pseudo R2	0.05	0.01	0.18	0.14	0.17	0.2
LR chisq(k)	166	10.4	566	453	493	637
Observations	4299	4299	4299	4299	4299	4299

Absolute value of z statistics in parentheses

* significant at 5%; ** significant at 1%

*** medium to high tech industries include auto parts production, drugs and pharmaceuticals, and

Table 7: Multinomial specification of location choice

	Specification (1)		Specification (2)	
	High FDI States	Low FDI States	High FDI States	Low FDI States
Business age	-0.130 (0.102)	0.034 (0.104)	-0.230 (0.097)*	-0.059 (0.098)
Startup size	-0.181 (0.068)**	-0.357 (0.074)**	-0.168 (0.063)**	-0.324 (0.068)**
Entrepreneurial roots located	-0.793 (0.338)*	-2.221 (0.318)**	-0.738 (0.333)*	-2.232 (0.310)**
Food	0.396 (0.323)	1.070 (0.323)**		
Textiles	-0.172 (0.271)	0.411 (0.274)		
Garments	2.771 (0.527)**	2.728 (0.535)**		
Drugs, pharmaceuticals and chemicals	-0.802 (0.196)**	-0.802 (0.215)**		
Auto parts	-0.028 (0.289)	1.255 (0.273)**		
Constant	2.250 (0.458)**	3.202 (0.453)**	2.426 (0.435)**	3.688 (0.426)**
Observations	1405		1405	
Log likelihood	- 1297		-1392	
Chisquare (k)	349		159	
Pseudo Rsquared	0.12		0.05	

Standard errors in parentheses

* significant at 5%; ** significant at 1%

Table 8: Conditional Expectation Correction of TFP Equation : CLogit selection equation
Dependent variable Log TFP (adjusted for skills)

	O L S				Random Effects (GLS)			
	All		High FDI States	Low FDI States	All		High FDI States	Low FDI States
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Adjustment for market power	1.841 (49.61)**		2.004 (33.55)**	1.692 (33.85)**	1.520 (51.49)**		1.501 (29.40)**	1.634 (43.09)**
Foreign market access	-0.121 (9.99)**	-0.156 (9.89)**	-0.132 (6.14)**	-0.155 (8.51)**	-0.114 (6.11)**	-0.143 (5.92)**	-0.133 (4.06)**	-0.142 (4.80)**
Domestic market access	2.925 (7.20)**	3.199 (6.01)**	2.820 (3.15)**	1.536 (3.49)**	3.164 (4.79)**	3.504 (4.09)**	3.160 (2.23)*	1.606 (2.11)*
Own industry local size	-0.209 (1.19)	0.006 (0.03)	0.022 (0.06)	-0.519 (2.08)*	-0.147 (0.53)	0.027 (0.07)	0.040 (0.07)	-0.354 (0.82)
Constrained by labor regulation (%)	-0.030 (2.60)**	-0.033 (2.16)*	-0.009 (0.30)	-0.019 (1.20)	-0.039 (2.08)*	-0.043 (1.78)	-0.022 (0.51)	-0.028 (1.02)
Output lost to power outage (%)	-0.037 (1.07)	0.124 (2.77)**	-0.030 (0.45)	0.088 (1.63)	-0.029 (0.57)	0.100 (1.50)	-0.002 (0.02)	0.075 (0.90)
Business age	0.159 (10.12)**	0.186 (9.06)**	0.134 (5.07)**	0.130 (6.43)**	0.161 (6.90)**	0.200 (6.58)**	0.131 (3.41)**	0.155 (4.93)**
High FDI state	-0.482 (9.59)**	-0.672 (10.23)**			-0.478 (6.25)**	-0.653 (6.59)**		
Low FDI state	-0.533 (10.30)**	-0.671 (9.90)**			-0.561 (6.99)**	-0.697 (6.70)**		
Selectivity term(1,2)	-1.241 (1.17)	-1.389 (1.00)	-0.037 (0.54)		-0.245 (0.50)	-0.184 (0.28)	-0.007 (0.20)	
Selectivity term(1,3)	0.972 (1.30)	1.142 (1.17)	0.102 (1.47)		0.134 (0.39)	0.103 (0.22)	0.011 (0.33)	
Selectivity term (3,1)	-0.748 (0.85)	-0.737 (0.64)		0.111 (1.20)	-0.260 (0.64)	-0.164 (0.30)		0.027 (0.72)
Selectivity term (3,2)	1.072 (0.99)	1.158 (0.81)		-0.126 (1.75)	0.266 (0.53)	0.178 (0.26)		-0.057 (1.91)
Constant	-23.909 (9.17)**	-17.558 (5.14)**	-25.590 (4.54)**	-12.852 (4.40)**	-24.612 (5.87)**	-20.259 (3.72)**	-25.676 (2.90)**	-13.967 (2.82)**
Industry dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Yer dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3440	3440	1322	1525	3440	3440	1322	1525
Establishments					1427	1427	555	604
R-squared	0.60	0.32	0.62	0.64	0.6	0.31	0.6	0.62
Rho					0.89	0.88	0.87	0.92

Absolute value of t statistics in parentheses

* significant at 5%; ** significant at 1%

Table 9 Correction for selectivity bias in TFP equation: correction via reduced form estimation
Dependent variable Log TFP (adjusted for skills)

	Conditional logit selection						Multinomial logit selection			
	OLS		Random effects (GLS)				OLS		Random effects (GLS)	
	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Adjustment for market power	1.869 (49.69)**		1.525 (51.51)**		1.526 (51.46)**	1.519 (51.22)**	1.845 (51.22)**		1.545 (50.92)**	
Foreign market access	-0.090 (7.56)**	-0.115 (7.39)**	-0.080 (4.35)**	-0.099 (4.13)**		-0.063 (3.64)**	-0.081 (7.06)**	-0.105 (6.87)**	-0.076 (4.27)**	-0.095 (4.03)**
Domestic market access	1.697 (4.41)**	1.738 (3.44)**	1.709 (2.74)**	1.771 (2.19)*		0.240 (0.41)	1.625 (4.36)**	1.610 (3.24)**	1.748 (2.89)**	1.828 (2.28)*
Own industry local size	0.216 (1.38)	0.445 (2.17)*	0.388 (1.59)	0.600 (1.89)		0.645 (2.74)**	-0.001 (0.01)	0.210 (1.06)	0.284 (1.21)	0.465 (1.50)
Constrained by labor regulation (%)	-0.032 (3.27)**	-0.046 (3.63)**	-0.039 (2.44)*	-0.053 (2.57)*	-0.009 (0.63)		-0.041 (4.47)**	-0.055 (4.48)**	-0.043 (2.85)**	-0.056 (2.81)**
Output lost to power outage (%)	-0.232 (7.83)**	-0.133 (3.43)**	-0.219 (4.83)**	-0.145 (2.47)*	-0.243 (6.04)**		-0.104 (3.53)**	0.007 (0.18)	-0.170 (3.87)**	-0.089 (1.53)
Business age	0.153 (9.64)**	0.182 (8.75)**	0.154 (6.55)**	0.194 (6.32)**	0.160 (6.74)**	0.159 (6.70)**	0.093 (5.38)**	0.105 (4.58)**	0.120 (5.04)**	0.151 (4.80)**
Estimated Pr(High FDI)	-0.099 (0.87)	-0.048 (1.04)	-0.066 (2.07)*	-0.080 (1.11)	-0.077 (2.09)*	-0.075 (1.18)	-3.021 (9.60)**	-3.801 (9.07)**	-1.172 (4.59)**	-1.448 (4.26)**
Estimated Pr(Low FDI)	-0.159 (2.00)*	-0.108 (0.32)	-0.077 (1.24)	-0.058 (1.16)	-0.064 (1.20)	-0.063 (2.03)*	-2.410 (13.20)**	-2.798 (11.50)**	-0.791 (5.32)**	-0.947 (4.79)**
Constant	-17.585 (6.93)**	-9.817 (2.95)**	-16.867 (4.11)**	-10.846 (2.03)*	-3.620 (19.59)**	-8.429 (2.17)*	-14.064 (5.70)**	-5.420 (1.65)	-16.069 (4.03)**	-9.739 (1.84)
Industry dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3440	3440	3440	3440	3440	3440	3440	3440	3440	3440
Establishments			1427	1427	1427	1427			1427	1427
R-squared	0.59	0.29	0.58	0.30	0.57	0.57	0.62	0.32	0.61	0.32
Rho			0.89	0.88	0.89	0.89			0.87	0.87
Hausman-wu test:										
df			15	14	12	13				
chisquare			44.71	42.53	25.16	76.61				

Absolute value of t statistics in parentheses

* significant at 5%; ** significant at 1%

Figure 2: Geographic sources of TFP gaps: selected cities, garments industry

Figure 3. Market access as a source of TFP gaps : selected cities , garments industry

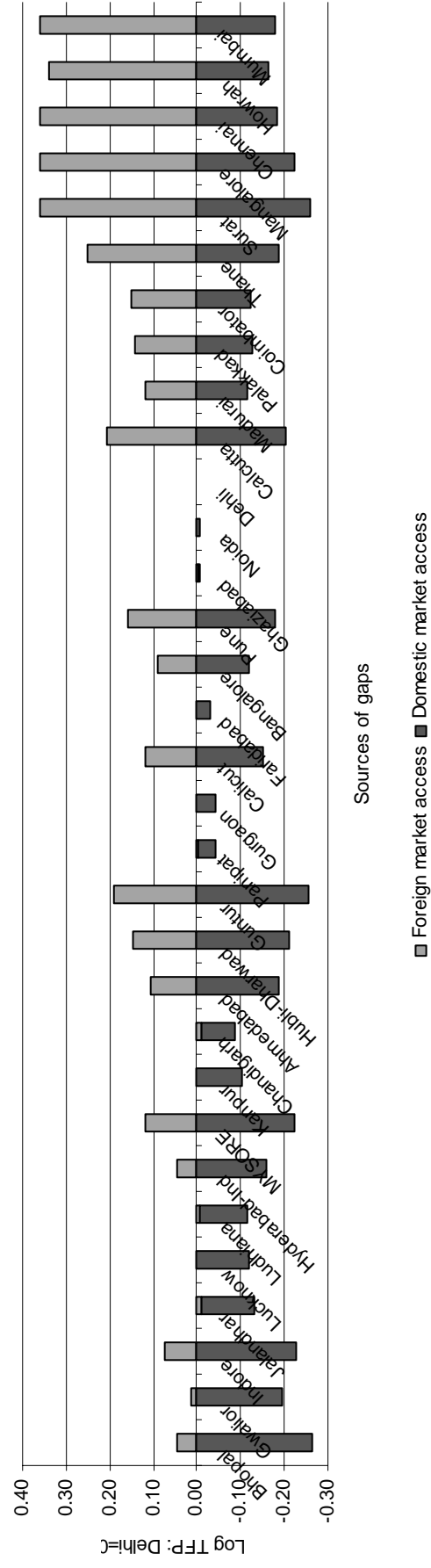


Figure 4: TFP gaps between selected cities: garments industry

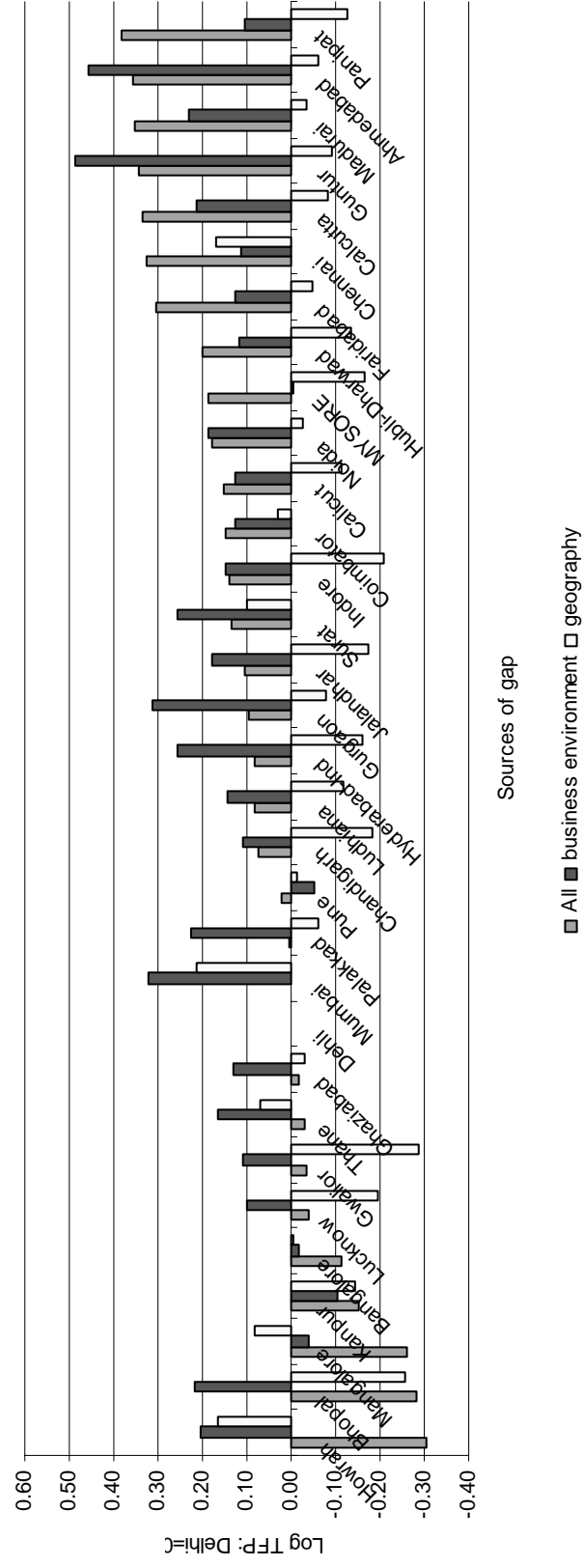


Figure 5. Business environmental sources of TFP gaps: selected cities, garments industry

